Modal Propositional Logic.

- **Propositional Logic:** Prop. Propositional variables p_i, $\wedge$, $\vee$, $\neg$, $\rightarrow$.
- **Modal Logic.** $\mathrm{Prop}+\ \Box$, $\Diamond$.
- **First-order logic.** $\mathrm{Prop}+\ \forall$, $\exists$, function symbols $\dot{\mathrm{f}}$, relation symbols $\dot{\mathrm{R}}$.

$$\mathbf{Prop} \quad \subseteq \quad \mathbf{Mod} \quad \subseteq \quad \mathbf{FOL}$$

Standard Translation

The standard translation (1).

Let $\dot{\mathrm{P}}_i$ be a unary relation symbol and $\dot{\mathrm{R}}$ a binary relation symbol.

We translate $\mathbf{Mod}$ into $\mathcal{L} = \{\dot{\mathrm{P}}_i, \dot{\mathrm{R}}\,;\, i \in \mathbb{N}\}$.

For a variable x, we define ST_x recursively:

$$\begin{aligned}
\mathrm{ST}_x(\mathrm{p}_i) &:= \dot{\mathrm{P}}_i(x) \\
\mathrm{ST}_x(\neg\varphi) &:= \neg\mathrm{ST}_x(\varphi) \\
\mathrm{ST}_x(\varphi \vee \psi) &:= \mathrm{ST}_x(\varphi) \vee \mathrm{ST}_x(\psi) \\
\mathrm{ST}_x(\Diamond\varphi) &:= \exists y \left(\dot{\mathrm{R}}(x,y) \wedge \mathrm{ST}_y(\varphi)\right)
\end{aligned}$$

The standard translation (2).

If $\langle M, R, V\rangle$ is a Kripke model, let $P_i := V(\mathrm{p}_i)$. If P_i is a unary relation on M, let $V(\mathrm{p}_i) := P_i$.

Theorem.

$$\langle M, R, V\rangle \models \varphi \quad \leftrightarrow \quad \langle M, P_i, R\,;\, i \in \mathbb{N}\rangle \models \forall x\, \mathrm{ST}_x(\varphi)$$

Corollary. Modal logic satisfies the compactness theorem.

Proof. Let Φ be a set of modal sentences such that every finite set has a model. Look at $\Phi^* := \{\forall x\, \mathrm{ST}_x(\varphi)\,;\, \varphi \in \Phi\}$. By the theorem, every finite subset of Φ^* has a model. By compactness for first-order logic, Φ^* has a model. But then Φ has a model. q.e.d.

Bisimulations.

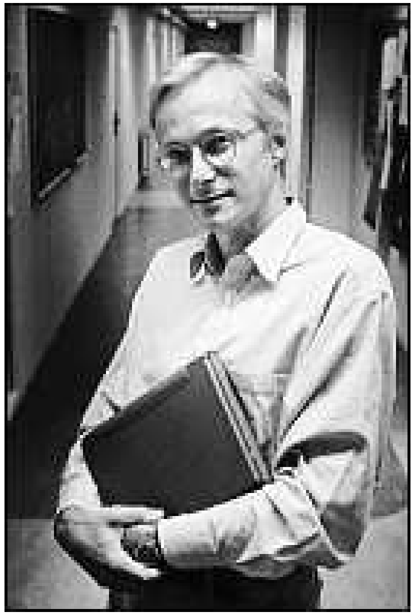
If $\langle M, R, V\rangle$ and $\langle M^*, R^*, V^*\rangle$ are Kripke models, then a relation $Z \subseteq M \times N$ is a bisimulation if

- If xZx^*, then $x \in V(\mathrm{p}_i)$ if and only if $x^* \in V(\mathrm{p}_i)$.
- If xZx^* and xRy, then there is some y^* such that $x^*R^*y^*$ and yZy^*.
- If xZx^* and $x^*R^*y^*$, then there is some y such that xRy and yZy^*.

A formula $\varphi(v)$ is called invariant under bisimulations if for all Kripke models $\mathbf{M}$ and $\mathbf{N}$, all $x \in M$ and $y \in N$, and all bisimulations Z such that xZy, we have

$$\mathbf{M} \models \varphi(x) \leftrightarrow \mathbf{N} \models \varphi(y).$$

van Benthem.


Johan van Benthem

Theorem (van Benthem; 1976). A formula in one free variable v is invariant under bisimulations if and only if it is equivalent to $\mathrm{ST}_v(\psi)$ for some modal formula ψ.

Modal Logic is the bisimulation-invariant fragment of first-order logic.

Intuitionistic Logic (1).

Recall the game semantics of intuitionistic propositional logic: $\models_{\text{dialog}} \varphi$.

- $\models_{\text{dialog}} \mathrm{p} \rightarrow \neg\neg \mathrm{p}$,
- $\not\models_{\text{dialog}} \neg\neg \mathrm{p} \rightarrow \mathrm{p}$,
- $\not\models_{\text{dialog}} \varphi \vee \neg\varphi$.

Kripke translation (1965) of intuitionistic propositional logic into modal logic:

$$\begin{aligned} \mathrm{K}(\mathrm{p}_i) &:= \Box \mathrm{p}_i \\ \mathrm{K}(\varphi \vee \psi) &:= \mathrm{K}(\varphi) \vee \mathrm{K}(\psi) \\ \mathrm{K}(\neg\varphi) &:= \Box\neg \mathrm{K}(\varphi) \end{aligned}$$

Intuitionistic Logic (2).

Theorem.

$$\models_{\text{dialog}} \varphi \leftrightarrow \mathbf{S4} \vdash \mathrm{K}(\varphi).$$

Consequently, φ is intuitionistically valid if and only if $\mathrm{K}(\varphi)$ holds on all transitive and reflexive frames.

$$\begin{array}{rcl}
\models_{\text{dialog}} \mathrm{p} \rightarrow \neg\neg\mathrm{p} & \rightsquigarrow & \Box\mathrm{p} \rightarrow \Box\Diamond\Box\mathrm{p} \\
\not\models_{\text{dialog}} \neg\neg\mathrm{p} \rightarrow \mathrm{p} & \rightsquigarrow & \Box\Diamond\Box\mathrm{p} \rightarrow \Box\mathrm{p} \\
\not\models_{\text{dialog}} \varphi \vee \neg\varphi & \rightsquigarrow & \mathrm{K}(\varphi) \vee \Box\neg\mathrm{K}(\varphi) \\
 & & \Box\mathrm{p} \vee \Box\neg\Box\mathrm{p} \\
 & & \Box\mathrm{p} \vee \Box\Diamond\neg\mathrm{p}
\end{array}$$

Provability Logic (1).

Leon Henkin (1952). “If φ is equivalent to PA $\vdash \varphi$, what do we know about φ?”

M. H. Löb, Solution of a problem of Leon Henkin, **Journal of Symbolic Logic** 20 (1955), p.115-118:

PA $\vdash ((\text{PA} \vdash \varphi) \rightarrow \varphi)$ implies PA $\vdash \varphi$.

Interpret $\Box\varphi$ as PA $\vdash \varphi$. Then Löb’s theorem becomes:

$$\text{(Löb)}\ \Box(\Box\varphi \rightarrow \varphi) \rightarrow \Box\varphi.$$

$\mathbf{GL}$ is the modal logic with the axiom (Löb).

Provability Logic (2).

Theorem (Segerberg-de Jongh-Kripke; 1971). $\mathbf{GL} \vdash \varphi$ if and only if φ is true on all transitive converse wellfounded frames.

A translation R from the language of model logic into the language of arithmetic is called a realization if

$$\begin{aligned} R(\bot) &= \bot \\ R(\neg\varphi) &= \neg R(\varphi) \\ R(\varphi \vee \psi) &= R(\varphi) \vee R(\psi) \\ R(\Box\varphi) &= \text{PA} \vdash R(\varphi). \end{aligned}$$

Theorem (Solovay; 1976). $\mathbf{GL} \vdash \varphi$ if and only if for all realizations R, PA $\vdash R(\varphi)$.

Modal Logics of Models (1).

One example: Modal logic of forcing extensions.

Joel D. Hamkins

A function H is called a Hamkins translation if

$$
\begin{aligned}
H(\bot) &= \bot \\
H(\neg\varphi) &= \neg H(\varphi) \\
H(\varphi \vee \psi) &= H(\varphi) \vee H(\psi) \\
H(\Diamond\varphi) &= \text{"there is a forcing extension in which } H(\varphi) \text{ holds".}
\end{aligned}
$$

The Modal Logic of Forcing: $\mathbf{Force} := \{\varphi\,;\ \mathsf{ZFC} \vdash H(\varphi)\}$.

Modal Logics of Models (2).

$\mathbf{Force} := \{\varphi\,;\, \mathsf{ZFC} \vdash H(\varphi)\}$.

Theorem (Hamkins).

1. $\mathbf{Force} \nvdash \mathbf{S5}$.
2. $\mathbf{Force} \vdash \mathbf{S4}$.
3. There is a model of set theory $\mathbf{V}$ such that the Hamkins translation of $\mathrm{S5}$ holds in that model.

Joel D. **Hamkins**, A simple maximality principle, **Journal of Symbolic Logic** 68 (2003), p. 527–550

Theorem (Hamkins-L). $\mathbf{Force} = \mathbf{S4.2}$.

Joel D. **Hamkins**, Benedikt **Löwe**, The Modal Logic of Forcing, **Transactions of the AMS** 360 (2008)

Tarski (1).

Alfred Tarski
1902-1983

- *Teitelbaum* (until c. 1923).
- 1918-1924. Studies in Warsaw. Student of Lesniewski.
- 1924. Banach-Tarski paradox.
- 1924-1939. Work in Poland.
- 1933. *The concept of truth in formalized languages*.
- From 1942 at the University of California at Berkeley.

Tarski (2).

- **Undefinability of Truth.**
- **Algebraic Logic.**
- **Logic and Geometry.**
  - A theory T admits elimination of quantifiers if every first-order formula is T-equivalent to a quantifier-free formula (Skolem, 1919).
  - **1955.** Quantifier elimination for the theory of real numbers ("real-closed fields").
  - Basic ideas of modern algebraic model theory.
  - Connections to theoretical computer science: running time of the quantifier elimination algorithms.

The puzzle of truth.

- Eubulides. “A man says he is lying. Is what he says true or false?”
- Sophismata.
- Buridan’s Proof of God’s Existence.

  (1) God exists.

  (2) (1) and (2) are false.

Tarski & Truth (1).

Alfred **Tarski**, The concept of truth in the languages of the deductive sciences, **Prace Towarzystwa Naukowego Warszawskiego, Wydzial III Nauk Matematyczno-Fizycznych** 34 (1933)

We say that a language $\mathcal{L}$ is saturated if there are

- an assignment $\varphi \mapsto \mathrm{t}_\varphi$ of $\mathcal{L}$-terms to $\mathcal{L}$-sentences,
- a surjective assignment $x \mapsto \mathrm{F}_x$ of $\mathcal{L}$-formulae in one free variable to objects.

Let T be an $\mathcal{L}$-theory and $\Phi(x)$ be an $\mathcal{L}$-formula with one free variable. We say that Φ is truth-adequate with respect to T if

- for all φ, either $T \vdash \Phi(\mathrm{t}_\varphi)$ or $T \vdash \neg\Phi(\mathrm{t}_\varphi)$ (totality), and
- for all φ, we have that

$$T \vdash \varphi \leftrightarrow \Phi(\mathrm{t}_\varphi)$$

(Adequacy; “Tarski’s T-convention”).

Tarski & Truth (2).

$$T \vdash \varphi \leftrightarrow \Phi(\mathrm{t}_\varphi).$$

Theorem (Undefinability of Truth). If $\mathcal{L}$ is saturated and T is a consistent $\mathcal{L}$-theory, then there is no formula Φ that is truth-adequate for T.

Proof. Suppose Φ is truth-adequate. Consider $\varphi(x) := \neg\Phi(\mathrm{t}_{\mathrm{F}_x(x)})$. This is a formula in one variable, there is some e such that $\mathrm{F}_e(x) = \neg\Phi(\mathrm{t}_{\mathrm{F}_x(x)})$. Consider $\mathrm{F}_e(e) = \neg\Phi(\mathrm{t}_{\mathrm{F}_e(e)})$.

$$\begin{array}{lll} T & \vdash & \mathrm{F}_e(e) \\ T & \vdash & \neg\Phi(\mathrm{t}_{\mathrm{F}_e(e)}) \\ T & \vdash & \neg\mathrm{F}_e(e) \qquad \text{(by adequacy)} \end{array}$$

So, Φ cannot be total. q.e.d.

Object language and metalanguage.

If $\mathcal{L}$ is any (interpreted) language, let $\mathcal{L}_{\mathbf{T}}$ be $\mathcal{L} \cup \{\mathbf{T}\}$ where $\mathbf{T}$ is a unary predicate symbol. If T is any consistent theory, just add the Tarski biconditional

$$\varphi \leftrightarrow \mathbf{T}(\mathrm{t}_\varphi)$$

to get $T_{\mathbf{T}}$.

Now $\mathbf{T}$ is a truth-adequate predicate with respect to $T_{\mathbf{T}}$, but only for sentences of $\mathcal{L}$.

The metalanguage $\mathcal{L}_{\mathbf{T}}$ can adequately talk about truth in the object language $\mathcal{L}$.

Unproblematic sentences.

- $\mathbf{T}(\mathrm{t}_{2+2=4})$. “$2+2=4$ is true.”
- $\mathbf{T}(\mathrm{t}_{\mathbf{T}(\mathrm{t}_{2+2=4})})$. “It is true that $2+2=4$ is true.”
- $\mathbf{T}(\mathrm{t}_{\neg\mathbf{T}(\mathrm{t}_{\mathbf{T}(\mathrm{t}_{2+2=4})})\rightarrow\varphi})$. “It is true that (If it is false that $2+2=4$ is true, then φ holds.)”

Well-foundedness.

An inductive definition of truth (1).

Let $\mathcal{L}$ be a language without truth predicate. We shall add a partial truth predicate $\mathbf{T}$ to get $\mathcal{L}_{\mathbf{T}}$:
Suppose we already have a partial truth predicate T interpreting $\mathbf{T}$. Then we can define $T^{+} := \{\mathrm{t}_{\varphi}\,;\,\varphi$ is true if $\mathbf{T}$ is interpreted by T.
Let

$$\begin{aligned} T_0 &:= \{\mathrm{t}_{\varphi}\,;\,\varphi \text{ is a true } \mathcal{L}\text{-sentence}\} \\ T_{i+1} &:= (T_i)^{+} \\ T_{\infty} &:= \bigcup_{i\in\mathbb{N}} T_i \end{aligned}$$

Then T_{∞} is a partial truth predicate that covers all of the “unproblematic” cases. All?

An inductive definition of truth (2).

$$
\begin{aligned}
T_0 &:= \{\mathrm{t}_\varphi \,;\, \varphi \text{ is a true } \mathcal{L}\text{-sentence}\} \\
T_{i+1} &:= (T_i)^+ \\
T_\infty &:= \bigcup_{i\in\mathbb{N}} T_i
\end{aligned}
$$

If φ is a formula, let $\mathbf{T}^0(\varphi) = \varphi$ and $\mathbf{T}^{n+1}(\varphi) = \mathbf{T}(\mathrm{t}_{\mathbf{T}^n(\varphi)})$.

Let ψ be the formalization of

> "For all n, $\mathbf{T}^n(2+2=4)$."

The formula ψ is not in the scope of any of the partial truth predicates T_i, so it can't be in T_∞.

But $\mathbf{T}(\mathrm{t}_\psi)$ is intuitively "unproblematic".

An inductive definition of truth (3).

More formally: T_∞ is not a fixed-point of the $^+$ operation.

$$T_\infty \subsetneq (T_\infty)^+.$$

Use ordinals as indices:

$$\begin{aligned} T_\omega &:= T_\infty \\ T_{\alpha+1} &:= (T_\alpha)^+ \\ T_\lambda &:= \bigcup_{\alpha<\lambda} T_\alpha \end{aligned}$$

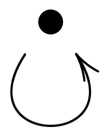
Theorem. There is a (countable) ordinal α such that $T_\alpha = T_{\alpha+1}$.

The source of the problem.

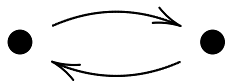
- What is the source of the problem with the Liar?
- Why didn't we have any problems with the "unproblematic" sentences?

Self-reference

- Liar. "This sentence is false."



- Nested Liar. "The second sentence is false."–"The first sentence is true."



- "This sentence has five words."

Pointer Semantics (1).

- Haim **Gaifman**, Pointers to truth, **Journal of Philosophy** 89 (1992), p. 223–261
- Haim **Gaifman**, Operational pointer semantics: solution to self-referential puzzles. I. Proceedings TARK II, p. 43–59
- Thomas **Bolander**, Logical Theories for Agent Introspection, PhD thesis, Technical University of Denmark 2003

Pointer Language: Let p_n be (countably many) propositional variables.

- Every $\mathbf{p}_n$ is an expression.
- $\perp$ and $\top$ are expressions.
- If E is an expression, then $\neg E$ is an expression.
- If E_i is an expression, then $\bigwedge_i E_i$ and $\bigvee_i E_i$ are expressions.

If E is an expression and n is a natural number, then $n : E$ is a clause. (**Interpretation.** "p_n states E".)

Pointer Semantics (2).

- Every $\mathbf{p}_n$ is an expression.
- $\bot$ and $\top$ are expressions.
- If E is an expression, then $\neg E$ is an expression.
- If E_i is an expression, then $\bigwedge_i E_i$ and $\bigvee_i E_i$ are expressions.

If E is an expression and n is a natural number, then $n : E$ is a clause.

Examples.

The Liar: $0 : \neg \mathrm{p}_0$.

The Truthteller: $0 : \mathrm{p}_0$.

One Nested Liar: $0 : \neg \mathrm{p}_1$.
$1 : \mathrm{p}_0$.

Two Nested Liars: $0 : \neg \mathrm{p}_1$.
$1 : \neg \mathrm{p}_0$.

Pointer Semantics (3).

- Every $\mathbf{p}_n$ is an expression.
- $\bot$ and $\top$ are expressions.
- If E is an expression, then $\neg E$ is an expression.
- If E_i is an expression, then $\bigwedge_i E_i$ and $\bigvee_i E_i$ are expressions.

If E is an expression and n is a natural number, then $n : E$ is a clause.

- An interpretation is a function $I : \mathbb{N} \to \{0, 1\}$ assigning truth values to propositional letters. I extends naturally to all expressions.
- If $n : E$ is a clause, we say that I respects $n : E$ if $I(n) = I(E)$.
- If Σ is a set of clauses, we say that it is paradoxical if there is no interpretation that respects all clauses in Σ.

Paradoxicality of the Liar.

The Liar:	$0 : \neg p_0.$	The Truthteller:	$0 : p_0.$
One Nested Liar:	$0 : \neg p_1.$ $1 : p_0.$	Two Nested Liars:	$0 : \neg p_1.$ $1 : \neg p_0.$

Paradoxical — Nonparadoxical

There are four relevant interpretations:

$$
\begin{array}{ll}
I_{00} & 0 \mapsto 0; 1 \mapsto 0 \\
I_{01} & 0 \mapsto 0; 1 \mapsto 1 \\
I_{10} & 0 \mapsto 1; 1 \mapsto 0 \\
I_{11} & 0 \mapsto 1; 1 \mapsto 1
\end{array}
$$

The Truthteller.

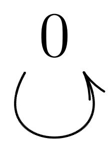
What is the problem with the truthteller and the two nested liars?
Both I_{01} and I_{10} are interpretations, so the two nested liars are nonparadoxical. But: the interpretations disagree about the truthvalues.
We call a set of clauses Σ determined if there is a unique interpretation.
The truthteller and the two nested liars are nonparadoxical but also nondetermined.

The dependency graph.

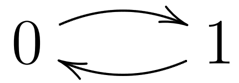
Let Σ be a (syntactically consistent) set of clauses. Then we can define the dependency graph of Σ as follows:

- $V := \{n\,;\, \mathrm{p}_n$ occurs in some clause in $\Sigma\}$.
- nEm if and only if $n : X \in \Sigma$ and p_m occurs in X.

Liar and Truthteller:



Nested Liar(s):



n is selfreferential if there is a path from n to n in the dependency graph.

Note. Selfreference does not imply paradoxicality!

Yablo's Paradox.

- Let $E_n := \bigwedge_{i>n} \neg \mathrm{p}_i$ and $\Upsilon := \{n : E_n \; ; \; n \in \mathbb{N}\}$.
- The dependency graph of Υ is $\langle \mathbb{N}, < \rangle$. No clause is self-referential in Υ.
- **Claim.** Σ is paradoxical.

  Proof. Let I be an interpretation.
  If $I(n) = 1$, then $\bigwedge_{i>n} \neg \mathrm{p}_i$ is true, so $I(i) = 0$ for all $i > n$, in particular for $i = n + 1$. But then $I(\bigwedge_{i>n+1} \neg \mathrm{p}_i) = 0$, so $I(\bigvee_{i>n+1} \mathrm{p}_i) = 1$. Pick i_0 such that $I(i_0) = 1$ to get a contradiction.

  So, $I(n) = 0$ for all n. But then $I(\bigwedge_n \neg \mathrm{p}_n) = 1$. Contradiction. q.e.d.

So: Paradoxicality does not imply self-reference.